REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY) 30-07-2015	2. REPORT TYPE Final Report	3. DATES COVERED (From - To) 1-Aug-2014 - 30-Apr-2015

4. TITLE AND SUBTITLE Final Report: A Novel Dataset of Structured Probability Distributions in Natural Scenes	5a. CONTRACT NUMBER W911NF-14-1-0489
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS Zhiyong Yang	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Medical College of Georgia Research Institu
1120 15th Street

Augusta, GA 30912 -4810

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
64575-LS-II.1

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

The study of natural scene statistics has served as a powerful framework for understanding vision and neural coding in the last several decades. Critical to this framework are datasets of natural scenes that have aligned multi-modal visual information, including luminance, color, stereoscopic disparity, movement, and three-dimensional (3D) information, which we are acquiring with support of DURIP grants from ARL/ARO and AFOSR. With support of this STIR grant, we performed statistical analyses on these datasets and developed a set of probabilistic models, referred to as probabilistic visual codes (PVCs). The PVCs are probabilistic models of static and dynamic, 2D and

15. SUBJECT TERMS
natural scene statistics, efficient codeing, independent components, probabilistic visual codes, 3D natural scenes

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Zhiyong Yang
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU	UU		19b. TELEPHONE NUMBER 706-721-4506

Report Title

Final Report: A Novel Dataset of Structured Probability Distributions in Natural Scenes

ABSTRACT

The study of natural scene statistics has served as a powerful framework for understanding vision and neural coding in the last several decades. Critical to this framework are datasets of natural scenes that have aligned multi-modal visual information, including luminance, color, stereoscopic disparity, movement, and three-dimensional (3D) information, which we are acquiring with support of DURIP grants from ARL/ARO and AFOSR. With support of this STIR grant, we performed statistical analyses on these datasets and developed a set of probabilistic models, referred to as probabilistic visual codes (PVCs). The PVCs are probabilistic models of static and dynamic, 2D and 3D natural scene patches in center-surround configurations. We found that these PVCs have a universal geometry: each PVC is a function of the total distance to hyperplanes in the spaces of 2D and/or 3D visual features in space and/or time domains and a large set of hyperplanes partition the feature spaces so that any natural scene patch is a combination of samples of PVCs. We are now exploring ways to relate PVCs to neural encoding and visual learning and applications of PVCs to visual saliency, natural 3D vision, scene vision, visual memory, object perception, and dynamic scene understanding.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received	Paper

TOTAL:

Number of Manuscripts:

Books

Received	Book

TOTAL:

Received	Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Not applicable.

Graduate Students

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Scientific progress and accomplishments

(1) Foreword

With support of this STIR grant, we made significant progress in analyzing the novel dataset we acquired recently. We developed a set of probabilistic visual codes (PVCs) and natural scene structures (NSSs). The PVCs are probabilistic models of static and dynamic, 2D and 3D natural scene patches in center-surround configurations. The NSSs, which can be encode by PVCs, provide a classification of natural scene patches. We examined the statistics of PVCs and NSSs and started to explore ways to relate PVCs to neural encoding and visual learning and applications of PVCs and NSSs to visual saliency, natural 3D vision, scene vision, visual learning, visual memory, object perception, and dynamic scene understanding.

We are in the process of preparing 3 manuscripts for journal submission.

(3) List of Appendixes, Illustrations and Tables

The attached PDF file contains the following figures and tables.

(4) Statement of the problem studied

Visual systems must inevitably adapt to the statistical characteristics of the natural environment [1-4]. Statistics of natural scenes have been used to account for many aspects of human natural vision [3]. The PI's works have rationalized many long-standing phenomena of brightness, color, geometrical forms, distance perception, and visual saliency [5-9]. Receptive fields (RFs) of simple and complex cells can be learned from natural scenes [10,11]; and the responses of V1 neurons in awake, behaving macaques suggest that classical and non-classical RFs form a sparse representation of the visual world [12,13]. More recently, neurons in the early visual cortex have been shown to adapt to image statistics very rapidly [14-16] and with great time precision at the population level [17], presumably via complex synaptic and network mechanisms [18].

We developed a dictionary of probabilistic visual codes (PVCs), i.e., probability distributions (PDs) of visual variables in static, dynamic, 2D, and 3D natural scenes.

1) A visual feature is a random variable and co-occurs at certain probabilities with other visual features in natural scenes.

2) These probabilities can be efficiently represented by a set of PDs in terms of independent components (ICs) of natural scenes [2].

3) The function of the visual system can be seen as encoding and operating on these PVCs to generate percepts that allow routine successful behaviors in the natural environment.

4) These PVCs can be used to achieve natural vision (e.g., visual saliency, scene vision, visual learning, and natural 3D vision).

(5) Summary of the most important results

Image registration

We model the alignment between any two images of the same scene acquired by our 2D and 3D image systems by a displacement, a rotation, a scale transform, and a non-linear mapping. For this purpose, we identify a number of features that are roughly uniformly distributed in the images and then find the best alignment (Fig. 1).

A dictionary of PVCs

We develop a dictionary of PVCs by modelling the joint PDs of natural scene patches that have aligned multi-modal visual information, including luminance, color, stereoscopic disparity, movement, and 3D information (Figs. 2-5). These PVCs suggest that there are joint classical-nonclassical RFs and joint 2D-3D RFs.

Universality of PVCs

There is a universal geometry in PVCs: each PVC is a function of the total distance to hyperplanes in the spaces of 2D and/or 3D visual features in space and/or time domains and a large set of hyperplanes partition the feature spaces so that any natural scene patch is a combination of samples of PVCs (Fig. 6).

PVCs convey bottom-up visual saliency

We can derive a measure of visual saliency from PVCs. Visual saliency is the perceptual quality that makes some items in visual scenes stand out from their immediate contexts [19]. Let P(Tm|Cnt) denote the maximal probability of a target T within context Cnt in natural scenes. We can define visual saliency SI of target T in context Cnt as

$SI = \log P(Tm|Cnt) - \log P(T|Cnt)$

Thus, SI of target T is a summation of the SI given by each of PVCs. This model of visual saliency is a good indicator of human gaze in free-viewing of static natural scenes (Table 1 and Fig. 7).

Encoding visual scenes by PVCs

We also propose natural scene structures (NSSs) as basic units of natural scenes. NSSs are patterns of co-occurrences of basic features in scene patches (~ a few degrees) and each NSS has a PD that describes its full range of natural variations (Figs. 10 and 11). A scene is a sample from a PD of visual scenes in terms of NSSs and their spatial arrangements. NSSs can be encoded by populations of PVCs.

Scaling law of NSSs

The frequency distribution of NSSs follows a power law (Fig. 12).

Fine correlational structures of NSSs

Co-occurrences of NSSs have very rich structures and have many patterns, most of which are very different from 1/f spectral [1-3] (on which many models of early visual processing are based) (Fig. 13).

NSSs and PVCs for scene categorization

NSSs can be used to encode natural scenes to achieve scene categorization (Table 2). This can developed into a new framework of scene vision.

NSSs and PVCs for visual learning

The fine structures of NSSs and PVCs provide ample opportunities for learning. Learning modifies the fine structures that improve visual discrimination (e.g., camouflage breaking) and recognition (Table 3 where the relative amplitudes of the ICs in the NSSs are modified by learning). This can developed into a new framework of visual learning.

3D vision based on NSSs and PVCs

For each NSS, we develop a set of joint 2D-3D RFs. To achieve natural 3D vision, we develop a hierarchical Bayesian inference framework based on NSSs and joint 2D-3D RFs (Fig. 14). We find that, in many situations, detailed, accurate 3D vision in natural conditions from a single monocular view is achievable based on NSSs and 2D-3D RFs (Fig. 16). This can be developed into a new framework of 3D vision.

(6) Bibliography

[1] Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation. Ann. Rev. Neurosci. 24, 1193-1216.

[2] Hyvärinen, A., Hurri, J. & Hoyer, P. O. (2009). Natural Image Statistics--A probabilistic approach to early computational vision. Springer.

[3] Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. Ann. Rev. Psy. 59, 167-192.

[4] Yang, Z. (2012). Vision as a fundamentally statistical machine. In: Visual Cortex – Current Status and Perspectives (S Molotchnikoff, ed.), 201-226. Intechopen.com.

[5] Yang, Z. & Purves, D. (2003). A statistical explanation of visual space. Nat. Neurosci. 6, 632-640.

[6] Yang, Z. & Purves, D. (2004). The statistical structure of natural light patterns determines perceived light intensity. Proc. Natl. Acad. Sci. USA 101, 8745-8750.

[7] Howe, C., Yang, Z. & Purves, D. (2005). The Poggendorff illusion explained by the statistics of natural scene geometry. Proc. Natl. Acad. Sci. USA 102,7707-7712.

[8] Long, F., Yang, Z. & Purves, D. (2006). Spectral statistics in natural scenes predict hue, saturation and brightness. Proc. Natl. Acad. Sci. USA 103, 6013-1018.

[9] Xu, J., Yang, Z. & Tsien, J. (2010). Emergence of visual saliency from natural scenes via context-mediated probability pistributions coding. PLoS ONE 5(12), e15796. doi:10.1371/journal.pone.0015796.

[10]Van Hateren, J. H. & Van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. Pro. Roy. Soc. Lond. 265, 359-66.

[11] Karklin, Y. & Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. Nature 457, 83-87.

[12] Vinje, W. E. & Gallant, J. L. (2000). Sparse coding and decorrelation in primary cortex during natural vision. Sci. 287, 1273-1276.

[13] Felsen, G., Touryan, J., Han, F. & Dan, Y. (2005). Cortical Sensitivity to Visual Features in Natural Scenes. PLoS Biol. 3 (10), e342. doi:10.1371/journal.pbio.0030342.

[14] Sharpee, T. et al. (2006). Adaptive filtering enhances information transmission in visual cortex. Nature 439, 936-942.

[15] Lesica, N. A. et al. (2007). Adaptation to stimulus contrast and correlations during natural visual stimulation. Neuron 55(3), 479-91.

[16] Gutnisky, D. A. & Dragoi, V. (2008). Adaptive coding of visual information in neural populations. Nature 452, 220-224.

[17] Butts, D. A. et al. (2007). Temporal precision in the neural code and the time scales of natural vision. Nature 449, 92-95.

[18] Haider, B. et al. (2010). Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. Neuron 65, 107-121.

[19] Tatler, B. W., Hayhoe, M. M., Land, M. F. & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. J.

Vision 11(5), 1–23.
[20] Itti, L, Koch, C. & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI 20, 1254-1259.
[21] Gao. D. & Vasconcelos, N. (2009). Decision-theoretic saliency: computational principles, biological plausibility, and implications for neurophysiology and psychophysics. Neural Comput. 21, 239-271.
[22] Zhang, L. et al. (2008). SUN: a Bayesian framework for saliency using natural statistics. J. Vision 8, 1-20.
[23] Bruce, N. D. B. & Tsotsos, J. K. (2009). Saliency, attention, and visual search: an information theoretic approach. J. Vision 9, 1-24.
[24] Gao, S., Tsang, I. W.-H. & Chia, L.-T. (2013). Laplacian Sparse Coding, Hypergraph Laplacian Sparse Coding, and Applications. IEEE Trans. PAMI. 35, 92-104.
[25] Dixit, M., Rasiwasia, N. & Vasconcelos, N. (2011) Adapted gaussian models for image classification. Proc. IEEE Comput. Soc. Conf. Comput. & Vis. Pattern Recognit. 1, 937-943.
[26] Wu, J. & Rehg, J. M. (2009). Beyond the euclidean distance: creating effective visual codebooks using the histogram intersection kernel. Proc. IEEE Intl. Conf. Comput. Vis. 1, 630-637.
[27] Wang, J., Yang, J., Yu, K., Lv, F. & Gong, Y. (2010). Locality-Constrained Linear Coding for Image Classification. Proc. IEEE Conf. Computer Vision and Pattern Recognition. 3360-3367.
[28] Li, L., Su, H., Xing, E.P. & Fei-Fei, Li. (2010). Object bank: a high-level image representation for scene classification and semantic feature sparsification. Proc. Adv. Neural Inf. Process. Syst. 22, 1378-1386.

Technology Transfer

Not applicable.

Scientific progress and accomplishments

(1) Foreword

With support of this STIR grant, we made significant progress in analyzing the novel dataset we acquired recently. We developed a set of probabilistic visual codes (PVCs) and natural scene structures (NSSs). The PVCs are probabilistic models of static and dynamic, 2D and 3D natural scene patches in center-surround configurations. The NSSs, which can be encode by PVCs, provide a classification of natural scene patches. We examined the statistics of PVCs and NSSs and started to explore ways to relate PVCs to neural encoding and visual learning and applications of PVCs and NSSs to visual saliency, natural 3D vision, scene vision, visual learning, visual memory, object perception, and dynamic scene understanding.

We are in the process of preparing 3 manuscripts for journal submission.

(3) List of Appendixes, Illustrations and Tables

The attached PDF file contains the following figures and tables.

(4) Statement of the problem studied

Visual systems must inevitably adapt to the statistical characteristics of the natural environment [1-4]. Statistics of natural scenes have been used to account for many aspects of human natural vision [3]. The PI's works have rationalized many long-standing phenomena of brightness, color, geometrical forms, distance perception, and visual saliency [5-9]. Receptive fields (RFs) of simple and complex cells can be learned from natural scenes [10,11]; and the responses of V1 neurons in awake, behaving macaques suggest that classical and non-classical RFs form a sparse representation of the visual world [12,13]. More recently, neurons in the early visual cortex have been shown to adapt to image statistics very rapidly [14-16] and with great time precision at the population level [17], presumably via complex synaptic and network mechanisms [18].

We developed a dictionary of probabilistic visual codes (**PVCs**), i.e., probability distributions (PDs) of visual variables in static, dynamic, 2D, and 3D natural scenes.

1) A visual feature is a random variable and co-occurs at certain probabilities with other visual features in natural scenes.

2) These probabilities can be efficiently represented by a set of PDs in terms of independent components (ICs) of natural scenes [2].

3) The function of the visual system can be seen as encoding and operating on these PVCs to generate percepts that allow routine successful behaviors in the natural environment.

4) These PVCs can be used to achieve natural vision (e.g., visual saliency, scene vision, visual learning, and natural 3D vision).

(5) Summary of the most important results

Image registration

We model the alignment between any two images of the same scene acquired by our 2D and 3D image systems by a displacement, a rotation, a scale transform, and a non-linear mapping. For this purpose, we identify a number of features that are roughly uniformly distributed in the images and then find the best alignment (**Fig. 1**).

A dictionary of PVCs

We develop a dictionary of PVCs by modelling the joint PDs of natural scene patches that have aligned multi-modal visual information, including luminance, color, stereoscopic disparity, movement, and 3D information (**Figs. 2-5**). These PVCs suggest that there are joint classical-nonclassical RFs and joint 2D-3D RFs.

Universality of PVCs

There is a universal geometry in PVCs: each PVC is a function of the total distance to hyperplanes in the spaces of 2D and/or 3D visual features in space and/or time domains and a large set of hyperplanes partition the feature spaces so that any natural scene patch is a combination of samples of PVCs (**Fig. 6**).

PVCs convey bottom-up visual saliency

We can derive a measure of visual saliency from PVCs. Visual saliency is the perceptual quality that makes some items in visual scenes stand out from their immediate contexts [19]. Let

P(Tm|Cnt) denote the maximal probability of a target T within context Cnt in natural scenes. We can define visual saliency SI of target T in context Cnt as

SI = log P(Tm|Cnt) - log P(T|Cnt)

Thus, SI of target T is a summation of the SI given by each of PVCs. This model of visual saliency is a good indicator of human gaze in free-viewing of static natural scenes (Table 1 and **Fig. 7**).

Encoding visual scenes by PVCs

We also propose natural scene structures (**NSSs**) as basic units of natural scenes. NSSs are patterns of co-occurrences of basic features in scene patches (~ a few degrees) and each NSS has a PD that describes its full range of natural variations (**Figs. 10** and **11**). A scene is a sample from a PD of visual scenes in terms of NSSs and their spatial arrangements. NSSs can be encoded by populations of PVCs.

Scaling law of NSSs

The frequency distribution of NSSs follows a power law (**Fig. 12**).

Fine correlational structures of NSSs

Co-occurrences of NSSs have very rich structures and have many patterns, most of which are very different from 1/f spectral [1-3] (on which many models of early visual processing are based) (**Fig. 13**).

NSSs and PVCs for scene categorization

NSSs can be used to encode natural scenes to achieve scene categorization (Table 2). This can developed into a new framework of scene vision.

NSSs and PVCs for visual learning

The fine structures of NSSs and PVCs provide ample opportunities for learning. Learning modifies the fine structures that improve visual discrimination (e.g., camouflage breaking) and recognition (Table 3 where the relative amplitudes of the ICs in the NSSs are modified by learning). This can developed into a new framework of visual learning.

3D vision based on NSSs and PVCs

For each NSS, we develop a set of joint 2D-3D RFs. To achieve natural 3D vision, we develop a hierarchical Bayesian inference framework based on NSSs and joint 2D-3D RFs (**Fig. 14**). We find that, in many situations, detailed, accurate 3D vision in natural conditions from a single monocular view is achievable based on NSSs and 2D-3D RFs (**Fig. 16**). This can be developed into a new framework of 3D vision.

(6) Bibliography

[1] Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Ann. Rev. Neurosci.* **24**, 1193-1216.

[2] Hyvärinen, A., Hurri, J. & Hoyer, P. O. (2009). Natural Image Statistics--A probabilistic approach to early computational vision. Springer.

[3] Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Ann. Rev. Psy*. **59**, 167-192.

[4] Yang, Z. (2012). Vision as a fundamentally statistical machine. In: Visual Cortex – Current Status and Perspectives (S Molotchnikoff, ed.), 201-226. Intechopen.com.

[5] Yang, Z. & Purves, D. (2003). A statistical explanation of visual space. *Nat. Neurosci.* **6**, 632-640.

[6] Yang, Z. & Purves, D. (2004). The statistical structure of natural light patterns determines perceived light intensity. *Proc. Natl. Acad. Sci. USA* **101**, 8745-8750.

[7] Howe, C., Yang, Z. & Purves, D. (2005). The Poggendorff illusion explained by the statistics of natural scene geometry. *Proc. Natl. Acad. Sci. USA* **102**,7707-7712.

[8] Long, F., Yang, Z. & Purves, D. (2006). Spectral statistics in natural scenes predict hue, saturation and brightness. *Proc. Natl. Acad. Sci. USA* **103**, 6013-1018.

[9] Xu, J., Yang, Z. & Tsien, J. (2010). Emergence of visual saliency from natural scenes via context-mediated probability pistributions coding. *PLoS ONE* **5(12)**, e15796. doi:10.1371/journal.pone.0015796.

[10]Van Hateren, J. H. & Van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Pro. Roy. Soc. Lond.* **265**, 359-66.

[11] Karklin, Y. & Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* **457**, 83-87.

[12] Vinje, W. E. & Gallant, J. L. (2000). Sparse coding and decorrelation in primary cortex during natural vision. *Sci.* **287**, 1273-1276.

[13] Felsen, G., Touryan, J., Han, F. & Dan, Y. (2005). Cortical Sensitivity to Visual Features in Natural Scenes. *PLoS Biol.* **3(10),** e342. doi:10.1371/journal.pbio.0030342.

[14] Sharpee, T. et al. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature* **439**, 936-942.

[15] Lesica, N. A. et al. (2007). Adaptation to stimulus contrast and correlations during natural visual stimulation. *Neuron* **55(3)**, 479-91.

[16] Gutnisky, D. A. & Dragoi, V. (2008). Adaptive coding of visual information in neural populations. *Nature* **452**, 220-224.

[17] Butts, D. A. et al. (2007). Temporal precision in the neural code and the time scales of natural vision. *Nature* **449**, 92-95.

[18] Haider, B. et al. (2010). Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron* **65**, 107-121.

[19] Tatler, B. W., Hayhoe, M. M., Land, M. F. & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *J. Vision* **11(5)**, 1–23.

[20] Itti, L, Koch, C. & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE PAMI* **20**, 1254-1259.

[21] Gao. D. & Vasconcelos, N. (2009). Decision-theoretic saliency: computational principles,

biological plausibility, and implications for neurophysiology and psychophysics. *Neural Comput.* **21**, 239-271.

[22] Zhang, L. et al. (2008). SUN: a Bayesian framework for saliency using natural statistics. *J. Vision* **8**, 1-20.

[23] Bruce, N. D. B. & Tsotsos, J. K. (2009). Saliency, attention, and visual search: an information theoretic approach. *J. Vision* **9**, 1-24.

[24] Gao, S., Tsang, I. W.-H. & Chia, L.-T. (2013). Laplacian Sparse Coding, Hypergraph Laplacian Sparse Coding, and Applications. *IEEE Trans. PAMI.* **35**, 92-104.

[25] Dixit, M., Rasiwasia, N. & Vasconcelos, N. (2011) Adapted gaussian models for image classification. *Proc. IEEE Comput. Soc. Conf. Comput. & Vis. Pattern Recognit.* **1**, 937-943.

[26] Wu, J. & Rehg, J. M. (2009). Beyond the euclidean distance: creating effective visual codebooks using the histogram intersection kernel. *Proc. IEEE Intl. Conf. Comput. Vis.* **1**, 630-637.

[27] Wang, J., Yang, J., Yu, K., Lv, F. & Gong, Y. (2010). Locality-Constrained Linear Coding for Image Classification. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. 3360-3367.

[28] Li, L., Su, H., Xing, E.P. & Fei-Fei, Li. (2010). Object bank: a high-level image representation for scene classification and semantic feature sparsification. *Proc. Adv. Neural Inf. Process. Syst.* **22**, 1378-1386.

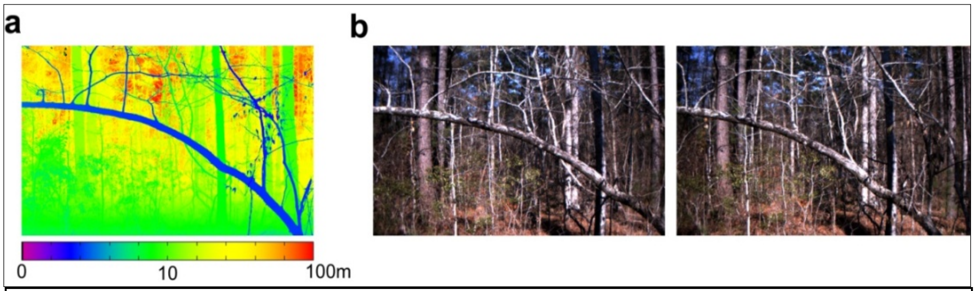


Fig. 1. Image registration. (**a**), Range image of a natural scene; the distance is indicated by color-coding. (**b**), A pair of stereoscopic images of the same scene.

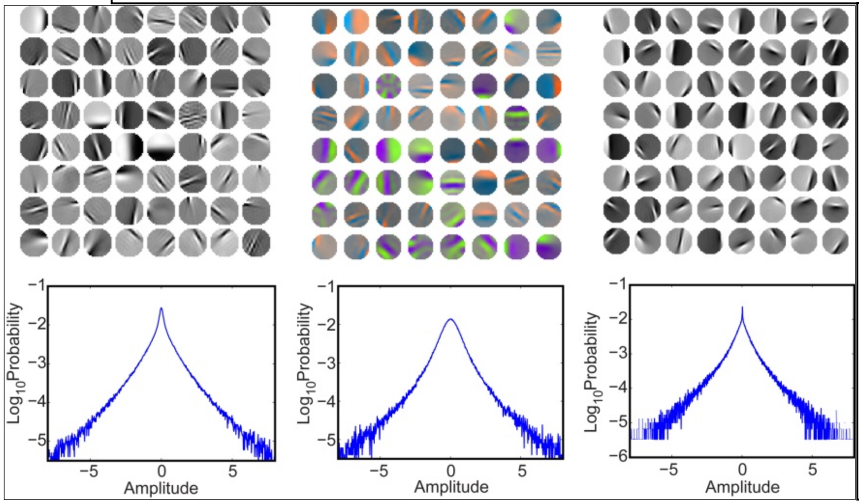


Fig. 2. ICs of natural scenes. Top row: 64 ICs. Bottom row: PD of the amplitude of an IC. Left column: luminance images of natural scenes. Middle column: color images of natural scenes. Right column: range images of natural scenes (bright/dark: far/near distance). These PDs are generalized Gaussian PDs (~exp(-**λ**|**x**|$^{\Delta}$)). For luminance and color images, Δ=0.5-3.5. For range images, Δ=0.5-1.

A visual feature is a random variable and co-occurs at certain probabilities with other visual features in natural scenes, denoted as **P(T|Cnt)** (**T:** target; **Cnt:** context). To obtain **P(T|Cnt)**, we model the PD of the 2D/3D scenes in the center given the scene patches in the center and the 6 surrounding circles as shown in **Fig. 3a**. The scene patch in the center serves as **T** and scene patches in the circles serve as **Cnt** in **P(T|Cnt)**. We use these notations: **[IC$_{2D}$]**: ICs of 2D image patches in the center; **[IC$_{X}$]**: ICs of 2D image patches in the 6 surrounding circles; **[IC$_{3D}$]**: ICs of range patches in the center; **Im**: 2D image; **R$_{m}$**: 3D image; ⊗: filtering by the filters of the ICs; and **W$_{C}$, W$_{X}$, V$_{2D}$, V$_{3D}$**: weight vectors.

P([IC$_{2D}$],[IC$_{X}$]) and joint C-NonC RFs. We represent P([IC$_{2D}$],[IC$_{X}$]) as a product of a set of PDs, each of which will have the form

$$\mathbf{(1/\Omega)exp(-\,\lambda|W_{C}[IC_{2D}\otimes I_{m}]+\,W_{X}[IC_{X}\otimes I_{m}]|^{\alpha}),}$$

where **Ω**, **λ**, and α are positive constants. Each is called a joint classical-nonclassical receptive field (**C-NonC RF**). 20 C-NonC RFs are shown in **Fig. 3c**.

We use this approach to obtain several other sets of PVCs, including 2D-3D RFs (**Fig. 4**), binocular 2D-3D RFs (**Fig. 5a**), spatial-temporal C-NonC RFs (Sp-Tm C-NonC RFs) (**Fig. 5b**), Sp-Tm 2D-3D RFs (**Fig. 5c**), and binocular Sp-Tm C-NonC RFs and Sp-Tm 2D-3D RFs.

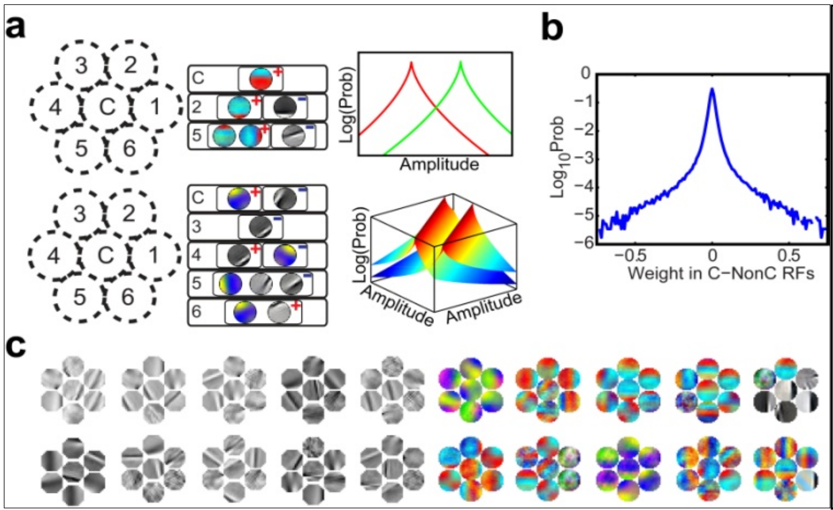


Fig. 3. C-NonC RFs. (**a**), Two C-NonC RFs. Left: Configuration of RFs (C: target; 1-6: context). Middle: Excitatory (+) and suppressive (-) component(s) (i.e., **ICs**). Right: PDs of $[IC_{2D}]$ for two sets of $[IC_X]$. (**b**), PD of the weights in C-NonC RFs. (**c**), 20 C-NonC RFs. In **Figs. 3-6**, and **7**, there are many ICs in each circle and only a few ICs with the greatest absolute weights are shown.

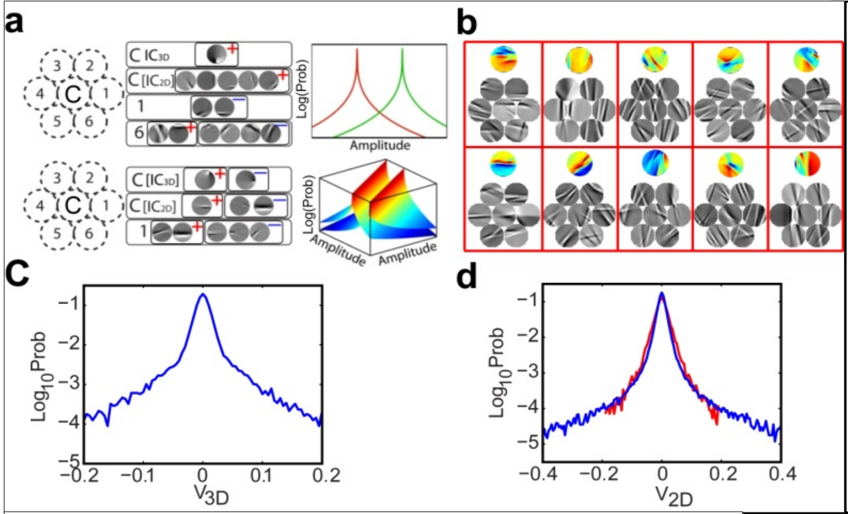


Fig. 4. 2D-3D RFs. (**a**), Two joint 2D-3D RFs. Left: Configuration of RFs (C: target; 1-6: context). Middle: Excitatory (+) and suppressive (-) component(s). Right: PDs of $[IC_{3D}]$ for two sets of values of $[IC_{2D}\ IC_X]$. (**b**), 10 2D-3D RFs. The colored patches are $[IC_{3D}]$ (red: far; blue: near). (**c**), PD of $\mathbf{V_{3D}}$. (**d**), PD of $\mathbf{V_{2D}}$. Red: PD of V_{2D} for $[IC_{2D}]$; blue: PD of V_{2D} for $[IC_X]$.

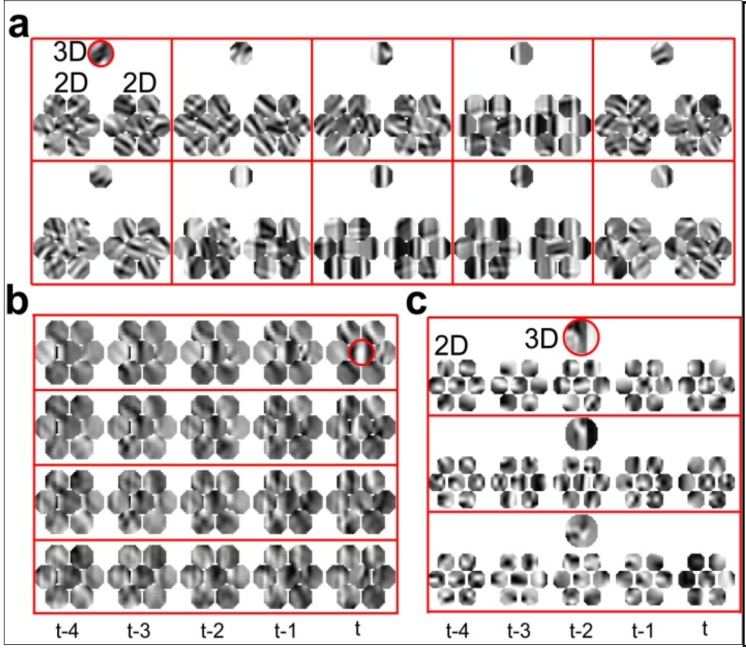


Fig. 5. Binocular 2D-3D RFs, Sp-Tm C-NonC RFs, and Sp-Tm 2D-3D RFs. (**a**), 10 binocular 2D-3D RFs. (**b**), 5 Sp-Tm C-NonC RFs. (**c**), 3 Sp-Tm 2D-3D RFs. In (**a**)-(**c**), each block shows an RF, the red circle is the target, all the other circles are the context, and t is time.

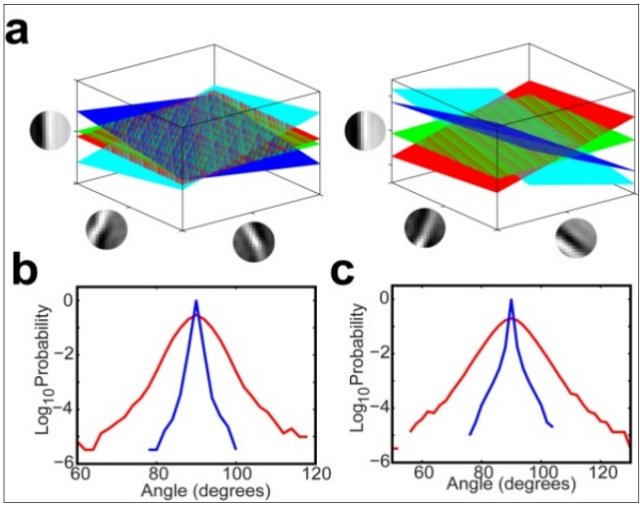
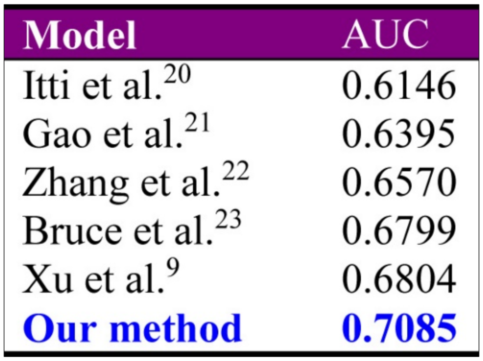


Fig. 6. Geometry of PVCs. (**a**), 4 hyperplanes in the space of 3D features (left) are paired with 4 hyperplanes in the space of 2D features (right) in a monocular 2D-3DRF. The paired hyperplanes have the same color and are plotted along three **ICs**. (**b**), PDs of the angles between two hyperplanes in C-NonC RFs. Red: target feature space. Blue: context's feature space. (**c**), PDs of the angles between two hyperplanes in monocular 2D-3D RFs. Red: space of 3D features. Blue: space of 2D features. Most hyperplanes are perpendicular to each other.



Model	AUC
Itti et al.[20]	0.6146
Gao et al.[21]	0.6395
Zhang et al.[22]	0.6570
Bruce et al.[23]	0.6799
Xu et al.[9]	0.6804
Our method	**0.7085**

Table 1. Performance of models of visual saliency. AUC: the area under the receiver operating characteristic curve formed by predicting fixations based on saliency. These results are based on ~10,000 fixations in 120 static color natural scenes[23].

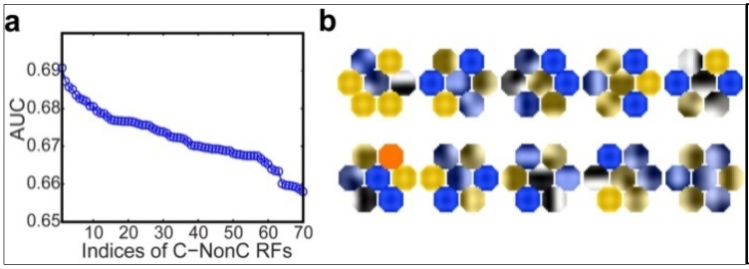


Fig. 7. Visual saliency predicted by C-NonC RFs. (**a**), AUC predicted by individual C-NonC RFs. (**b**), The first 10 C-NonC RFs that have the highest AUC shown in (**a**).

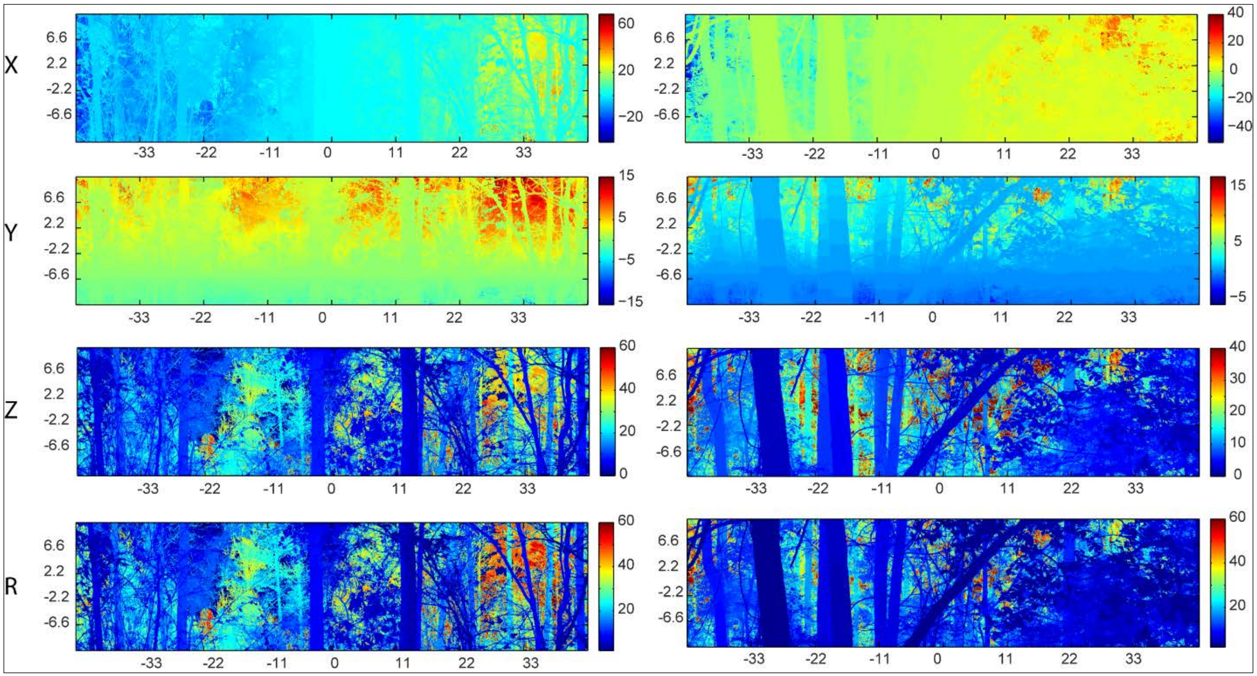


Fig. 8. Examples of 3D natural scenes. The X, Y, Z, and R components in meters are shown in color coding for two scenes (left and right). The horizontal axis is azimuthal angle (in degrees) and the horizontal axis is polar angle.

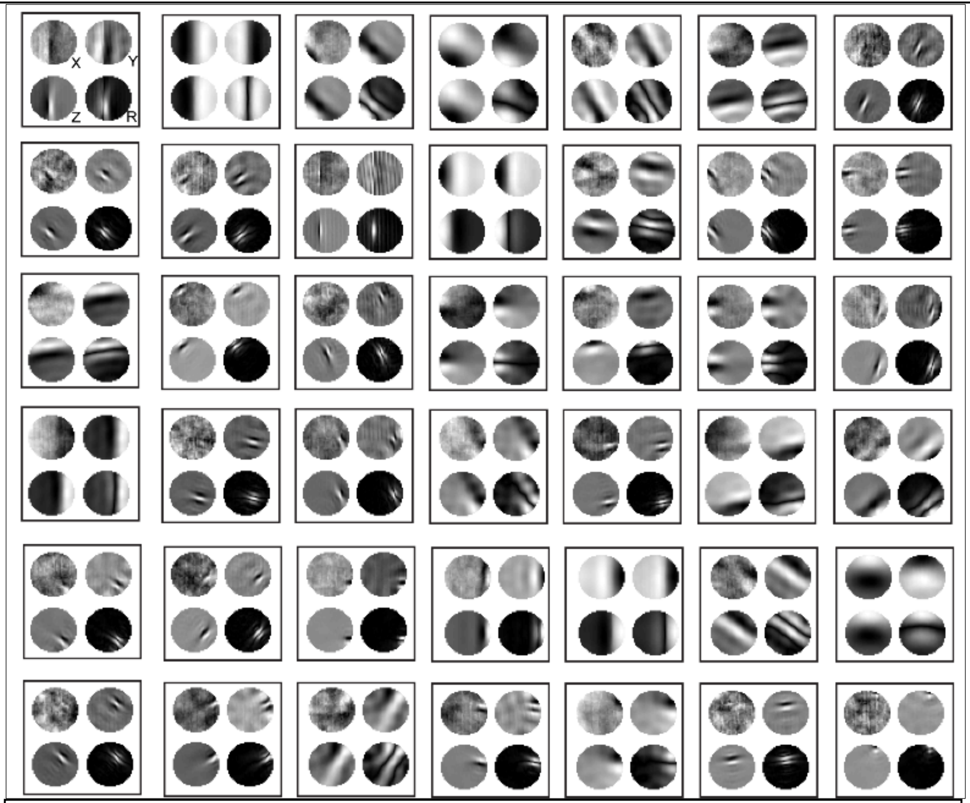


Fig. 9. ICs of 3D natural scenes. The X, Y, Z, and R components are indicated in the first panel.

Five steps to compile natural scene structures (NSSs) (**Fig. 10**).

1. Sample a large number of circular patches in a hexagon configuration at multiple spatial scales.

2. Perform independent component analysis on the circular patches and obtain ICs at each spatial scale.

3. Fit Gabor functions to the ICs and classify the ICs at multiple spatial scales into a set of clusters (referred to as IC clusters) using the parameters of the fitted Gabor functions as features.

4. Map the circular patches to the IC clusters, compute the features of the circular patches, and pool the features of the patches in the hexagon configuration at multiple spatial scales.

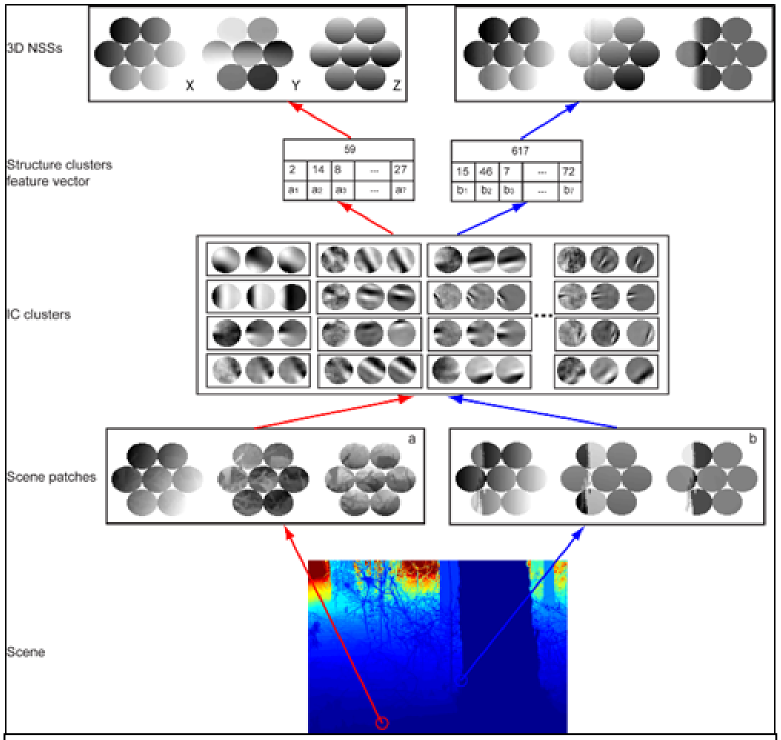


Fig. 10. Procedure for compiling NSSs. One spatial scale is shown here.

5. Partition the space of feature vectors into a set of NSSs.

In contrast to simple features, NSSs are highly structured mid-level representations that are building blocks of natural scenes. Each of the NSSs contains patches of natural scenes that entails a specific pattern of concatenations of local features in natural scenes. Note that each of the NSSs shown in **Fig. 11** is the average of a large number of scene patches that share the same structure.

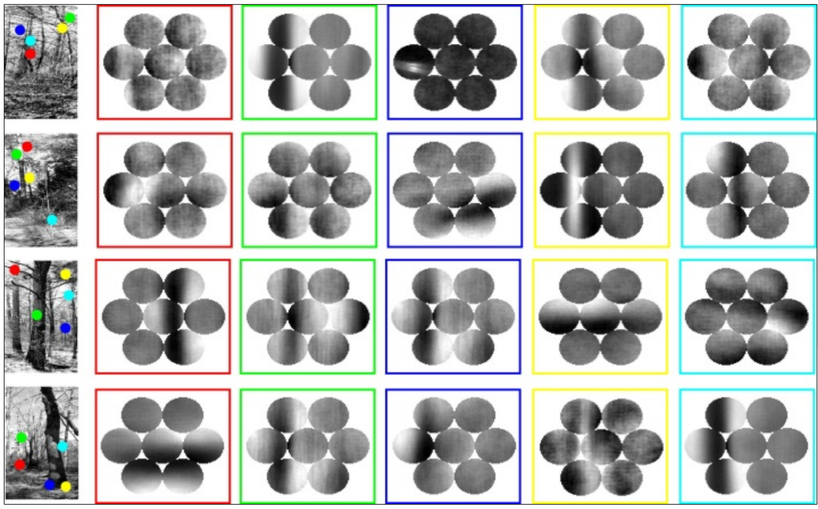


Fig. 11. Examples of NSSs. Five NSSs compiled from each of the 4 selected scenes. The locations of the NSSs in the scenes and the boxes of NSSs are indicated by the same color.

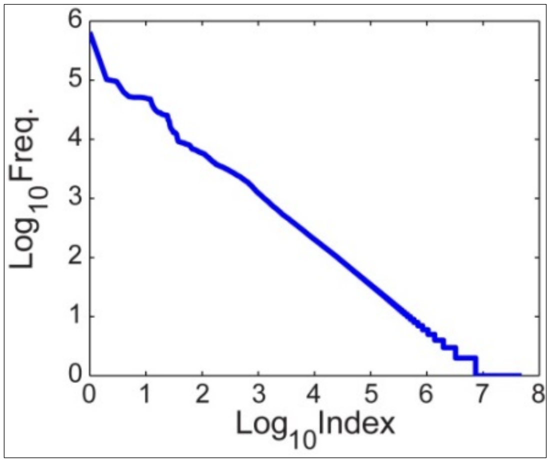


Fig. 12. Distribution of frequencies of NSSs. The indices of NSSs are ordered based on the occurring frequencies.

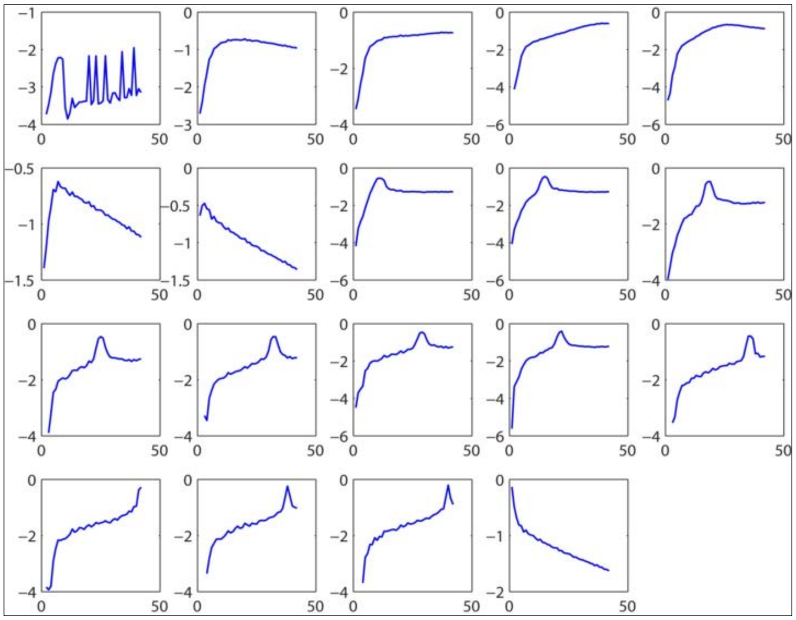
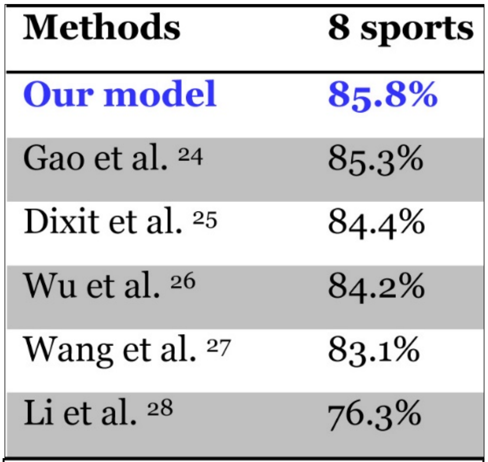
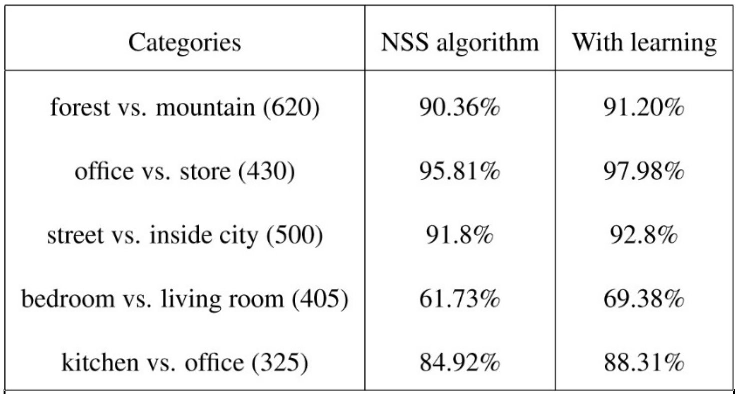


Fig. 13. Patterns of co-occurrences of pairs of NSSs. X-axis: spatial separation in degrees of visual angle; Y-axis: logarithm (base=10) of normalized numbers of co-occurrences.



Methods	8 sports
Our model	85.8%
Gao et al. [24]	85.3%
Dixit et al. [25]	84.4%
Wu et al. [26]	84.2%
Wang et al. [27]	83.1%
Li et al. [28]	76.3%

Table 2. Performance of a model based on NSSs and other models on scene categorization. The dataset contains scenes of 8 sports.



Categories	NSS algorithm	With learning
forest vs. mountain (620)	90.36%	91.20%
office vs. store (430)	95.81%	97.98%
street vs. inside city (500)	91.8%	92.8%
bedroom vs. living room (405)	61.73%	69.38%
kitchen vs. office (325)	84.92%	88.31%

Table 3. **Effect of learning on scene categorization**. The numbers in the parentheses are the numbers of images being tested. 50 images per category are used to train the model.

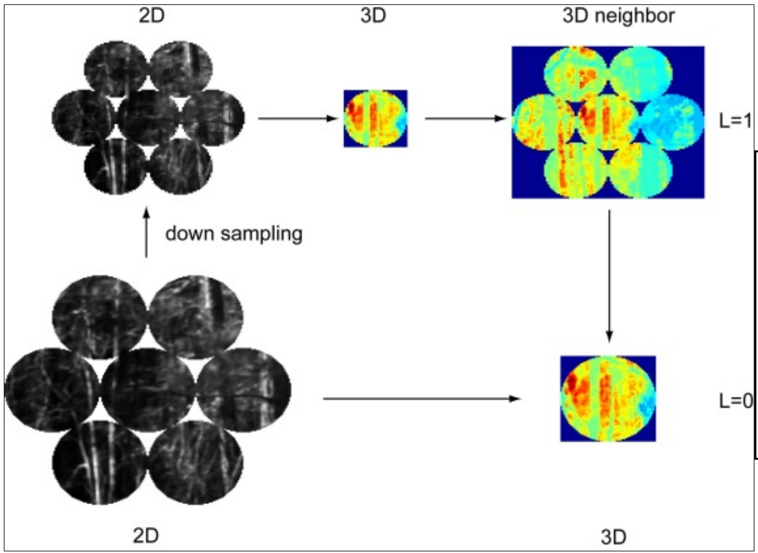


Fig. 14. Pyramid representation and hierarchical Bayesian inference. Both 2D and 3D scenes are represented by pyramids and 3D scenes underlying 2D scenes can be estimated from hierarchical 2D-3D RFs. The arrow indicates the flow of inference.

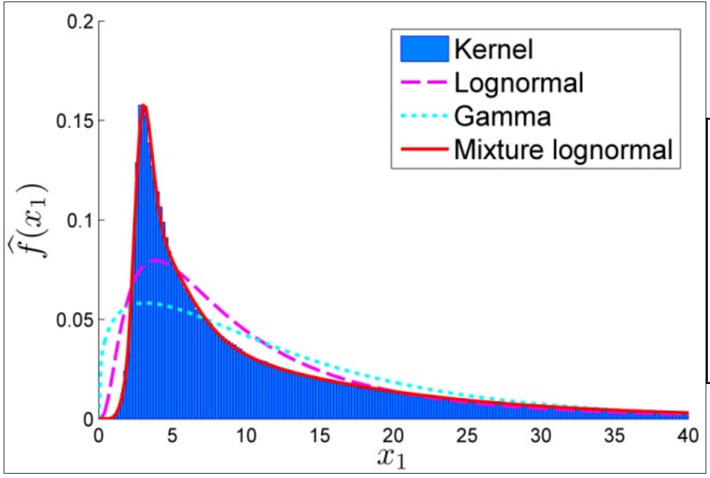


Fig. 15. PD of distances in 3D natural scenes. Probability densities of distances (meters) estimated by four methods, a mixture of log-normal distributions, a log-normal distribution, a gamma distribution, and kernel density estimation.

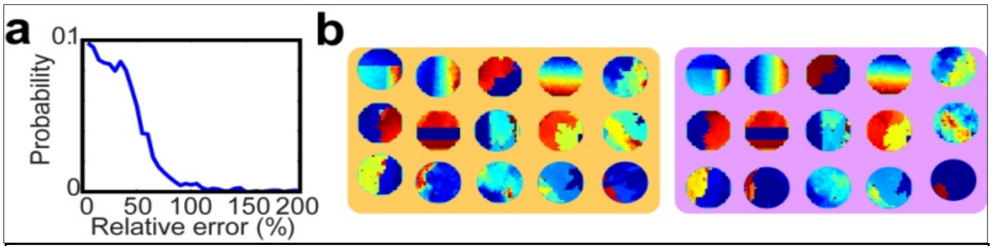


Fig. 16. Natural 3D vision based on 2D-3D RFs. (**a**), PD of 3D reconstruction error. (**b**). Examples of reconstructed (left panel) and ground-truth 3D scene patches (right panel). Red/blue: far/near distance.